732A90 Computational Statistics
Lecture 6
Derivation of EM algorithm for Slide 15

Krzysztof Bartoszek

February 28, 2020

E–step: derive

$$Q(\theta, \theta^k) = \mathrm{E}\left[\mathrm{loglik}(\theta|Y,Z)|\theta^k, Y\right] = \mathrm{E}\left[-\frac{1}{2}n\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(y_i-\mu)^2\right)|\theta^k, Y\right] \overset{*}{=} .$$

Our observed data is $Y = \{y_1, \ldots, y_r\}$ and the latent variables (unobserved data) is $Z = \{y_{r+1}, \ldots, y_n\}$. Hence, continuing we have

$$\begin{aligned}
&\overset{*}{=} -\tfrac{1}{2}n\log(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}\mathrm{E}\left[\sum_{i=1}^{n} y_i^2 - 2\mu\sum_{i=1}^{n} y_i + n\mu^2|\theta^k, Y\right] \\
&= -\tfrac{1}{2}n\log(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}\mathrm{E}\left[\sum_{i=1}^{r} y_i^2 + \sum_{i=r+1}^{n} y_i^2 - 2\mu\sum_{i=1}^{r} y_i - 2\mu\sum_{i=r+1}^{n} y_i + n\mu^2|\theta^k, Y\right] \\
&= -\tfrac{1}{2}n\log(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}\sum_{i=1}^{r} y_i^2 + \sum_{i=r+1}^{n}\mathrm{E}\left[y_i^2|\theta^k\right] - 2\mu\sum_{i=1}^{r} y_i - 2\mu\sum_{i=r+1}^{n}\mathrm{E}\left[y_i|\theta^k\right] + n\mu^2 \\
&= -\tfrac{1}{2}n\log(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}\sum_{i=1}^{r} y_i^2 + \sum_{i=r+1}^{n}(\mu_k + \sigma_k^2) - 2\mu\sum_{i=1}^{r} y_i - 2\mu\sum_{i=r+1}^{n}\mu_k + n\mu^2 \\
&= -\tfrac{1}{2}n\log(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}\sum_{i=1}^{r}(y_i-\mu)^2 + (n-r)(\mu-\mu_k)^2 + (n-r)\sigma_k^2 \\
&= \{\text{setting the unobserved } y_i \equiv \mu_k\} \\
&= -\tfrac{1}{2}n\log(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2 + \tfrac{n-r}{2\sigma^2}\sigma_k^2 = Q(\theta, \theta^k).
\end{aligned}$$

M–step: taking the derivatives with respect to the parameters μ and σ^2

$$\begin{aligned}
&Q(\theta, \theta^k)'_{\mu} = \tfrac{2}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu), \\
&Q(\theta, \theta^k)'_{\sigma^2} = -\tfrac{n}{2}\tfrac{1}{2\pi\sigma^2}\cdot 2\pi - \left(\sum_{i=1}^{n}(y_i-\mu)^2\right)\tfrac{-1}{2}\tfrac{1}{(\sigma^2)^2} - \tfrac{1}{2}(n-r)\sigma_k^2(-1)\tfrac{1}{(\sigma^2)^2}.
\end{aligned}$$

Setting the first equation to 0 we obtain $0 = \sum_{i=1}^{n}(y_i - \mu)$, giving $n\mu = \sum y_i$ and in turn

$$\mu_{k+1} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Turning to the second equation

$$\begin{aligned}
&0 = -\tfrac{n}{2}\tfrac{1}{2\pi\sigma^2}\cdot 2\pi - \left(\sum_{i=1}^{n}(y_i-\mu)^2\right)\tfrac{-1}{2}\tfrac{1}{(\sigma^2)^2} - \tfrac{1}{2}(n-r)\sigma_k^2(-1)\tfrac{1}{(\sigma^2)^2}\\
&0 = -\tfrac{n}{2} + \tfrac{1}{(\sigma^2)^2}\left(\sum_{i=1}^{n}(y_i-\mu)^2\right) + \tfrac{\sigma_k^2}{(\sigma^2)^2}(n-r)\\
&n\sigma^2 = \left(\sum_{i=1}^{n}(y_i-\mu)^2\right) + (n-r)\sigma_k^2\\
&\sigma^2 = \tfrac{1}{n}\left(\sum_{i=1}^{n}(y_i-\mu)^2\right) + \tfrac{1}{n}(n-r)\sigma_k^2\\
&= \tfrac{1}{n}\left(\sum_{i=1}^{n} y_i^2 - 2\mu\sum_{i=1}^{n} y_i + n\mu^2 + (n-r)\sigma_k^2\right)\\
&= \tfrac{1}{n}\left(\sum_{i=1}^{r} y_i^2 + \sum_{i=r+1}^{n}\mu_k^2 - 2\mu\sum_{i=1}^{r} y_i - 2\mu\sum_{i=r+1}^{n}\mu_k + n\mu^2 + (n-r)\sigma_k^2\right)\\
&= \tfrac{1}{n}\left(\sum_{i=1}^{r} y_i^2 + (n-r)\mu_k^2 - 2\mu\sum_{i=1}^{r} y_i - 2(n-r)\mu\mu_k + n\mu^2 + (n-r)\sigma_k^2\right)\\
&= \tfrac{1}{n}\left(\sum_{i=1}^{r} y_i^2 + (n-r)(\mu_k^2+\sigma_k^2)\right) + \mu^2 - \tfrac{2}{n}\mu\sum_{i=1}^{n} y_i.
\end{aligned}$$

As we now have the estimate at the $(k+1)$st step of μ as μ_{k+1}, then

$$\mu_{k+1}^2 - \frac{2}{n}\mu_{k+1}\sum_{i=1}^{n} y_i = \mu_{k+1}^2 - 2\mu_{k+1}\frac{1}{n}\sum_{i=1}^{n} y_i = \mu_{k+1}^2 - 2\mu_{k+1}^2 = -\mu_{k+1}^2$$

and we obtain

$$\sigma_{k+1}^2 = \frac{1}{n}\left(\sum_{i=1}^{r} y_i^2 + (n-r)(\mu_k^2\sigma_k^2)\right) - \mu_{k+1}^2.$$

The next step could then be to e.g. consider the second derivative and show that this is actually a maximum.